Research article

Identification of methylation driver genes for predicting the prognosis of pancreatic cancer patients based on whole-genome DNA methylation sequencing technology

Chao Song [a,b,c,1], Ganggang Wang [d,1], Mengmeng Liu [e,1], Zijin Xu [c], Xin Liang [f], Kai Ding [f], Yu Chen [f], Wenquan Wang [b,**], Wenhui Lou [b,***], Liang Liu [a,b,*]

[a] *Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai, 200000, China*
[b] *Department of Pancreatic Surgery, Affiliated Zhongshan Hospital of Fudan University, Shanghai, 200000, China*
[c] *Department of General Surgery, Affiliated Zhongshan Hospital of Fudan University, Qingpu Branch, Shanghai, 200000, China*
[d] *Department of Hepatobiliary Surgery, Pudong Hospital, Fudan University, Shanghai, 200000, China*
[e] *Department of Gastroenterology, Affiliated Zhongshan Hospital of Fudan University, Qingpu Branch, Shanghai, 200000, China*
[f] *CAS Key Laboratory of Nutrition, University of Chinese Academy of Sciences, Shanghai, 200000, China*

A R T I C L E   I N F O

A B S T R A C T

This study was based on the use of whole-genome DNA methylation sequencing technology to identify DNA methylation biomarkers in tumor tissue that can predict the prognosis of patients with pancreatic cancer (PCa). TCGA database was used to download PCa-related DNA methylation and transcriptome atlas data. Methylation driver genes (MDGs) were obtained using the MethylMix package. Candidate genes in the MDGs were screened for prognostic relevance to PCa patients by univariate Cox analysis, and a prognostic risk score model was constructed based on the key MDGs. ROC curve analysis was performed to assess the accuracy of the prognostic risk score model. The effects of PIK3C2B knockdown on malignant phenotypes of PCa cells were investigated *in vitro*. A total of 2737 differentially expressed genes were identified, with 649 upregulated and 2088 downregulated, using 178 PCa samples and 171 normal samples. MethylMix was employed to identify 71 methylation-driven genes (47 hypermethylated and 24 hypomethylated) from 185 TCGA PCa samples. Cox regression analyses identified eight key MDGs (LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5, PIK3C2B, IGF1R) associated with prognosis in PCa. Seven of them were hypermethylated, while PIK3C2B was hypomethylated. A prognostic risk prediction model was constructed based on the eight key MDGs, which was found to accurately predict the prognosis of PCa patients. In addition, the malignant phenotypes of PANC-1 cells were decreased after the knockdown of PIK3C2B. Therefore, the prognostic risk prediction model based on the eight key MDGs could accurately predict the prognosis of PCa patients.

 * Corresponding author. Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai, 200000, China.
 ** Corresponding author.
 *** Corresponding author.
 E-mail addresses: wang.wenquan@zs-hospital.sh.cn (W. Wang), lou.wenhui@zs-hospital.sh.cn (W. Lou), liuliang.zlhospital@fudan.edu.cn (L. Liu).
 [1] Co-first authors.

1. Introduction

Pancreatic cancer (PCa) is the 14th most common cancer in the world, and the 7th highest cause of cancer mortality, with a 5–7% 5-

---

A list of abbreviation words

DCA   Decision curve analysis
DEG   Differentially expressed gene
DMG   Differentially methylated gene
MDG   Methylation driver gene
MSK   Memorial Sloan-Kettering Cancer
NCI   National Cancer Institute
NHGRI   National Human Genome Research Institute
ROC,   Receiver Operating Characteristic Curve
TCGA   The Cancer Genome Atlas

---

year survival rate, and the age-standardized incidence of PCa increased by 1.03 % per year between 1973 and 2014, meaning that it is expected to rise from the 4th to the 2nd most common cause of cancer-related deaths in the US by 2030 [1]. Incidence rates vary widely by country, with age-standardized incidence rates highest in Europe and North America and lowest in Africa and South Central Asia, with a general trend towards higher incidence rates in developed countries compared with developing countries and higher incidence rates of PCa in both men and women in higher HDI countries [2]. The causes of PCa are complex, and the main causes that are relatively difficult to modify are age, sex, race, blood group, gut microbiota, diabetes, family history, and genetic predisposition [3]; the wide variation in the incidence of PCa between countries also suggests that environmental factors play an important role as risk factors for the disease and could be improved by modifying the environment and lifestyle. Examples include smoking, alcohol, chronic pancreatitis, obesity, dietary factors, and H. pylori infection [4].

Currently, analytical methods used for whole-genome DNA methylation sequencing are divided into (1) track-specific methods (MSP, methyl light, combined bisulfite restriction analysis, bisulfite-sanger sequencing, bisulfite pyrose sequencing, base-specific cleavage coupled to mass spectrometry, heavy methylation, methylation-sensitive single nucleotide primer extension, methylation-sensitive melting curve analysis, methylation-sensitive high-resolution melting); (2) global estimation of methylcytosine content (aminomethylation analysis); (3) genome-scale methods (restriction landmark genome scanning, methylation-specific arbitrary primer PCR, amplification of inter methylation sites, methylation-sensitive representation differential analysis); (4) microarray-based methods (restriction enzyme treatment, affinity enrichment, bisulfite treatment) [5]. DNA promoter hypermethylation as an early oncogenic mechanism, researchers developed a diagnostic prediction model based on an optimized accelerated bisulfite treatment protocol, methylation-specific PCR of 28 genes on plasma samples. They found that cell-free DNA promoter hypermethylation has the potential to be a diagnostic marker for pancreatic adenocarcinoma and to distinguish between malignant and benign pancreatic disease [6]. Using a whole-genome pharmacological transcriptomic approach to identify novel cancer-specific DNA methylation alterations in PCa cell lines, the investigators identified promoter DNA methylation of BNC1 and ADAMTS1 among eight promising genes as potential biomarkers for the detection of early PCa [7].

Important genes associated with cancer development are called driver genes, and driver genes determine the most important cause of this cancer. Methylated driver genes are oncogenes driven by DNA methylation. If a gene is under-expressed in a cancer sample compared to a normal sample, this gene is highly methylated in the cancer sample. Therefore, we consider this gene a driver gene, and its differential methylation value (DM-value) is its degree of differential methylation. For mixed samples of multiple identical cancer types, it is also possible to subtype that cancer type based on their DM-values [8]. A growing body of research suggests that altered DNA methylation plays an important role in PCa, with Koutsioumpa et al. finding that lysine (K)-specific methyltransferase 2D (KMT2D), which is subject to biallelic CpG methylation, undergoes transcriptional repression and gene inactivation in human pancreatic tumors thereby promoting tumor growth; in parallel, knockdown of KMT2D promotes tumor growth by regulating SLC2A3 increased aerobic glycolysis and proliferation rates and altered the lipidome profile of PCa cells [9]. Researchers identified eight genes in 65 invasive PCa samples that showed aberrant hypermethylation, such as the SARP2 gene. Epigenetic inactivation caused by aberrant methylation could confer resistance to apoptosis, leading to PCa progression [10].

Our study aimed to screen for DNA methylation markers that predict prognosis in PCa patients based on whole-genome DNA methylation sequencing technology. Scientifically, our study provides a strong theoretical basis for screening eight key MDGs (LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5, PIK3C2B, and IGF1R) that could be used as prognostic markers for PCa patients by whole-genome DNA methylation sequencing technology for the first time, which is of great scientific importance. PCa is a very lethal solid tumor with a poor prognosis. Surgery remains the basis for the cure of PC, but the majority (80–85 %) of patients presenting with unresectable or metastatic tumors are lost to surgery [11]. In addition to advances in surgical techniques and chemotherapeutic approaches, exploring effective and new biomarkers is another effective way to improve early diagnosis rates and predict prognosis in patients with PCa [12]. The prognostic risk prediction model constructed in our study based on eight key MDGs could accurately predict the prognosis of PCa patients for early preventive and therapeutic measures, which is expected to improve the survival rate of

PCa patients and bring hope for survival to PCa patients and is of great clinical importance.

2. Materials and methods

2.1. Acquisition and analysis of DNA methylation data associated with PCa

The Cancer Genome Atlas (TCGA) is a joint project of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) initiated in 2006. It is an important source of clinical data, genomic variation, mRNA expression, miRNA expression, methylation, and other data for cancer researchers.

We retrieved human PCa methylation data and transcriptomic data from the TCGA database. The methylation data consisted of 185 tumor samples and 10 adjacent non-tumor samples, obtained from the Illumina Human Methylation 450k platform. The transcriptomic analysis data included 178 tumor samples and 4 adjacent non-tumor samples, with no information on subtype expression and miRNA expression quantification. To augment our study, we also obtained a set of 167 pancreatic transcriptome sequencing data from the GTEx database, specifically from the normal control group. This data was downloaded from the UCSC-XENA website (http://xena.ucsc.edu/).

First, we normalized the downloaded data and used the limma package in R to obtain differentially expressed genes (DEGs) and differentially methylated genes (DMGs). Then, we used the MethylMix package for integrated analysis.

MethylMix is a statistical package for R software. The function of this package is to identify potential methylation driver oncogenes by integrating DNA methylation data and RNA expression data to identify methylation driver genes (MDGs), i.e., if a gene is highly expressed and hypomethylated, or low expressed and hypermethylated, and the gene expression level is significantly negatively correlated with the methylation level, then we consider this gene is considered to be MDGs.

Using the MethylMix algorithm, we calculated the correlation between gene methylation levels and gene expression and constructed β mixed models using $|logFC| > 0.5$, $P < 0.05$, and $Cor < −0.3$ to obtain MDGs [13].

The data provided by TCGA is publicly available and does not require the approval of the local ethics committee.

2.2. GEO database

In this study, we obtained the PCa RNA-seq microarray dataset (GSE183795) from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). The sequencing platform used was GPL6244, and the microarray dataset included a total of 139 PCa tissues and 102 adjacent tissues.

2.3. Functional enrichment analysis of GO and KEGG for DEGs

After obtaining MDGs, to further understand the biological functions of these selected genes, we performed GO analysis and KEGG pathway enrichment analysis using the "ClusterProfiler" R package for GO and KEGG enrichment analysis at the biological process (BP), cellular component (CC) and $P < 0.05$ was set as the standard threshold, and the results were visualized using Goplot [14].

2.4. Constructing a prognostic risk score model based on MDGs

To analyze the relationship between MDGs and prognosis, we performed a one-way Cox analysis using the "survival" package in conjunction with the prognostic outcome of PCa in TCGA. Lasso regression and multivariate Cox analysis were performed on the significant correlates (genes) from the univariate Cox analysis results. Risk values were calculated for each PCa patient to construct a prognostic risk score model. The "survival" package and the "survminer" package (https://CRAN.R-project.org/package=survivalsvm) were used to draw prognostic. The survival curve of the risk score model was calculated, and the p-value was calculated. *Patients were* divided into high-risk and low-risk groups according to the median risk score. Kaplan-Meier survival analysis was performed to compare the difference in overall survival (OS) between the high- and low-risk groups, with $P < 0.05$ selected as the cut-off value. ROC curve analysis was performed on the prognostic risk score model using the "survivalROC" package (https://CRAN.R-project.org/package=survivalROC) and the prognostic risk score model AUC (Area Under Curve) values [15,16].

2.5. Construction and validation of Nomograms

To expand the clinical application and usability of prognostic risk scoring models, a column line graph was constructed that could be easily used to calculate the expected survival of individual PCa patients by combining the genes involved in the construction of the prognostic risk scoring model with common clinicopathological features, using the 'rms' package to Nomograms were drawn for the genes involved in the construction of the prognostic risk score using the 'rms' package and calibration curves were drawn to assess the agreement between actual and predicted values. decision curve analysis (DCA) is an evaluation method developed by Memorial Sloan Kettering Cancer Institute (MSKCC) in 2006 [17]. DCA is a method for evaluating and comparing multiple clinical prediction models [18].

2.6. Analysis of genetic alterations

The cBio Cancer Genomics Portal, cancer genomics portal (cBioPortal), is a major online platform for analyzing cancer genomics data. cBio Cancer Genomics Portal (http://cbioportal.org), funded by Memorial Sloan-Kettering The cBio Cancer Genomics Portal (.), funded by Memorial Sloan-Kettering Cancer (MSK), is designed to address a large amount of data available from extensive sample tumor genomics studies to facilitate the more accessible and more direct application of large sample tumor genomic findings to oncology research [19].

LungCNS/Brain is a tool for visualizing cancer genomic data analysis. Genomics data analysis could be performed through a simple set-up of the web interface to explore the association between genetic alterations and clinicopathological features. No special bio-informatics knowledge is required. The results are presented graphically and easy to understand, and article-quality graphs could be produced to facilitate the use of the results. To date, 169 studies derived from multiple databases such as TCGA, ICGC, and published papers are included (data list: Head/Neck http://www.cbioportal.org/data_sets.jsp); data types include somatic mutations, DNA copy number alterations, mRNA and microRNA expression, DNA methylation, protein abundance; data available for download. We used cBioPortal to study genetic alterations in MDGs in PCa patients [20].

2.7. GSEA pathway enrichment analysis

Gene Set Enrichment Analysis (GSEA) is used to assess the trend in the distribution of genes in a predefined gene set in a table of genes ranked by their phenotypic relevance to determine their contribution to the phenotype. The input data consists of two parts, a gene set of known functions (either GO annotations, MsigDB annotations, or other formatted gene set definitions) and an expression matrix (also a sorted list), where the software sorts the genes according to their phenotypic relevance (understood as the change in expression value) from highest to lowest and then determines whether the genes under each annotation in the gene set are enriched in the upper or lower part of the gene list after the phenotypic correlation ranking, and thus determine the effect of synergistic changes in genes within this gene set on phenotypic changes. This is distinctly different from using significant differential genes for enrichment analysis, which screens for differential genes and then determines in which annotated pathways the differential genes are enriched; this involves setting a differential threshold, which is subjective and could only be used for genes with large variation in expression, i. e., what we define as significant differential genes. GSEA, on the other hand, is not limited to differential genes and is theoretically more likely to encompass the effects of subtle but coordinated changes in biological pathways, especially in sets of genes with less significant fold differences [21,22], from the perspective of gene set enrichment. The GSEA pathway enrichment analysis was performed using the "fgsea" package and the "clusterprofiler" of R software.

All analytical methods and R packages above were performed using v4.0.3 of R software (R Foundation for Statistical Computing, 2020). $p < 0.05$ was considered statistically significant.

2.8. Cell culture and transfection

The normal human pancreatic cell line HPDE6c7 was purchased from Shanghai Zeye Biotechnology Co., Ltd. (AC338285). The PANC-1 PCa cell line was obtained from the ATCC (CRL-1469). The PCa cell lines, BxPC-3 (CBP60542) and AsPC-1 (CBP60546), were purchased from Nanjing Cobioer Biotechnology Co., Ltd. All the aforementioned cell lines were cultured in DMEM (10569044, Gibco, USA) supplemented with 10 % FBS (100099141, Gibco, USA) and 1 % penicillin-streptomycin (15070063, Gibco, USA) and incubated in a CO_2 incubator at 37 °C [23].

The cells were digested with trypsin when PANC-1 and BxPC-3 cells were in the log phase, inoculated in 6-well plates at 1×10^5 cells per well, and incubated routinely for 24 h. After the cells had reached about 40 % fusion, the cells were infected according to the lentivirus infection instructions and divided into the sh-NC group (infected with the no-load control lentivirus sh-NC), sh-PIK3C2B-1 group (infected with the lentivirus sh -PIK3C2B-) and sh-PIK3C2B-2 (infected with the lentivirus sh-PIK3C2B-2), and cells were collected 72 h after transfection for subsequent experiments. shRNA sequences are shown in Table S1.

2.9. RT-qPCR

Total tissue or cellular RNA was extracted using Trizol (16096020, Thermo Fisher Scientific, New York, NY, USA). cDNA was obtained by reverse transcription using a reverse transcription kit (RR047A, Takara, Japan). cDNA was spiked using an SYBR Premix Ex TaqTM II kit (DRR081, Takara, Japan). Spiking was performed, and samples were subjected to RT-qPCR reactions in a real-time fluorescence quantitative PCR instrument (ABI 7500, ABI, Foster City, CA, USA). GAPDH was used as an internal reference for the coding gene. RT-qPCR was set up with 3 replicate wells. The 2-ΔΔCt method was used to quantify gene expression [24]. The primer design is shown in Table S2.

2.10. Western Blot analysis

Total protein was extracted from cells using RIPA lysis buffer (P0013B, Beyotime Biotechnology) and incubated on ice for 30 min. The lysate was then centrifuged at 8000 g for 10 min at 4 °C, and the supernatant was collected. Protein concentration was determined using the BCA assay kit (A53226, ThermoFisher, USA). After separation by polyacrylamide gel electrophoresis, the proteins were transferred to a PVDF membrane (IPVH85R, Millipore, Germany) using the wet transfer method. The membrane was blocked with 5 %

BSA at room temperature for 1 h and then incubated overnight at 4 °C with mouse anti (MA5-26243, Invitrogen) and rabbit anti-GAPDH (ab9485, Abcam) primary antibodies. After washing the membrane three times with TBST (each for 10 min), it was incubated for 1 h with HRP-conjugated goat anti-rabbit IgG H&L secondary antibody (ab97051, diluted 1:2000, Abcam, Cambridge, UK) or HRP-conjugated goat anti-mouse IgG H&L secondary antibody (31430, 1:2000, Invitrogen). After TBST washing, the membrane was placed on a clean glass plate. An appropriate amount of A and B solutions from the ECL fluorescence detection kit (abs920, Absin Biotechnology Co., Ltd., Shanghai, China) were mixed and then applied to the membrane in a darkroom. The relative protein levels were analyzed using Quantity One V4.6.2 software from Bio-Rad (USA), with the grayscale value of the corresponding protein band normalized to the grayscale value of the GAPDH protein band [25].

2.11. CCK-8

Cell viability after lentivirus treatment was assessed using the CCK-8 (K1018, Apexbio, USA) kit by spreading 1×10^4 cells per well in 96-well plates (100 μL/well), followed by the addition of 10 μL of CCK-8 solution at each time point (0 h, 24 h, 48 h, and 72 h) for 2 h at 37 °C. The absorbance of each sample at 480 nm was measured by an enzyme marker (Bio-Rad, Hercules, CA, USA) to measure the absorbance of each well at 480 nm [26].

2.12. Scratch assay

Cells were scraped from the bottom of the wells using a 10 μl tip, and the floating cells were gently washed twice using a DMEM medium. Images of each well were captured using a microscope (magnification, × 100) at 0 and 24 h intervals after wounding (Olympus X51; Olympus, Tokyo, Japan) [27].

2.13. Transwell assay

Transwell chambers with 8 μm pore size polycarbonate membranes containing matrigel were pre-filled with 600 mL of DMEM medium with 20 % FBS in the lower chamber and equilibrated at 37 °C for 1 h. PANC-1 and BxPC-3 cells were resuspended in DMEM medium without FBS after 48 h of transfection, and 1×10^6/mL of cells were inoculated into the upper chamber and incubated at 37 °C with 5 % CO_2 for 24 h. The cells were fixed with 4 % paraformaldehyde for 20 min and stained with 0.1 % crystal violet for 10 min. The
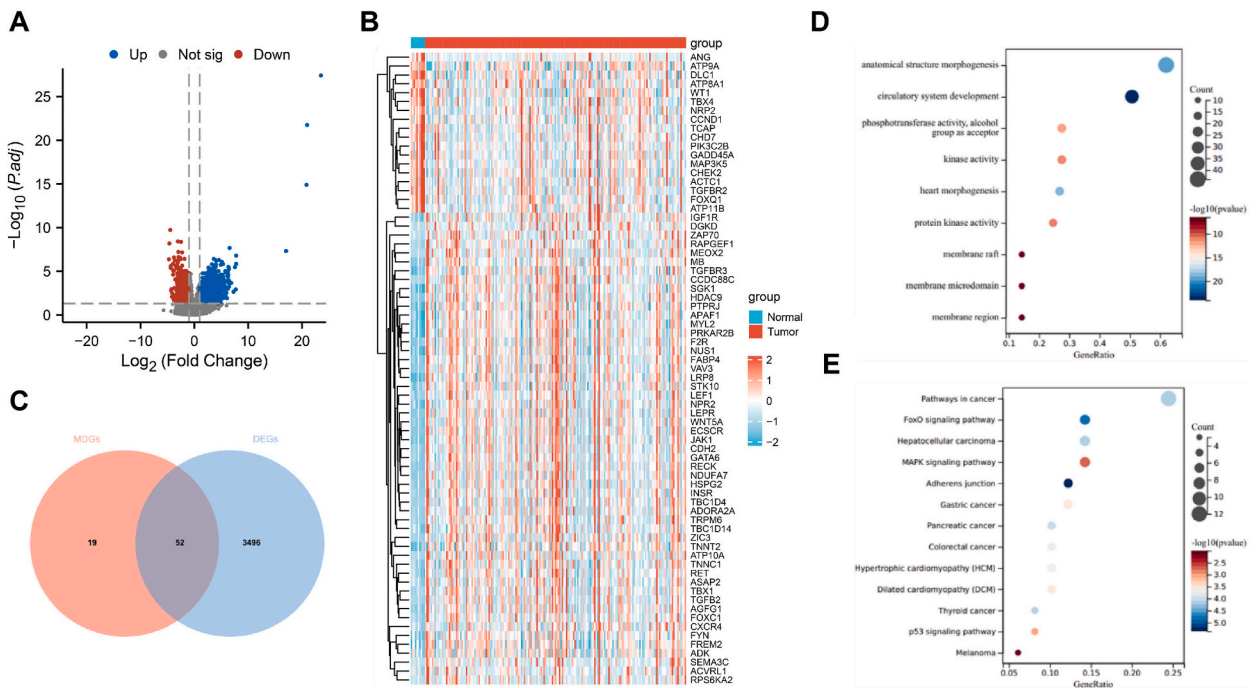

Fig. 1. Screening of DEGs and MDGs in PCa samples based on the TCGA database.
Note: A: Volcano plot showing DEGs between normal and PCa samples in the TCGA database (blue dots represent upregulated genes, red dots indicate downregulated genes). B: Heatmap analysis of MDGs in PCa from the TCGA database (where the colors range from green to red, indicating the trend from low to high methylation). C: Venn diagram showing the intersection between DEGs and MDGs. D: Bubble chart presenting the results of the GO analysis for MDGs. E: Bubble chart illustrating the results of the KEGG enrichment analysis for MDGs (where the size of the circles represents the number of selected genes and the color represents the P-value of the enrichment analysis).

surface cells were wiped off, and the cells were observed under an inverted fluorescent microscope (TE2000, Nikon, Japan). Five fields of view were photographed. The number of cells passing through the chambers was counted [26].

2.14. Statistical analysis

All statistical analyses of our study data were performed using SPSS 21.0 (IBM, USA) statistical software. Measures were expressed as mean ± standard deviation, first tested for normality and chi-square, tested for conformity to normal distribution and chi-square, and unpaired *t*-test between two groups; data were compared between groups at different time points using repeated measures ANOVA with Bonferroni method for post-hoc tests. $p < 0.05$ indicated statistically significant differences.

3. Results

3.1. Identification of 2737 DEGs and 71 MDGs in PCa samples through bioinformatics analysis

A total of 3548 DEGs were identified from 178 PCa samples and 171 normal samples (4 from TCGA database, 167 from GTEx dataset). Among these DEGs, 1752 were upregulated, and 1796 were downregulated (Fig. 1A). MDGs analysis using the MethylMix package identified 71 MDGs (50 hypermethylated and 21 hypomethylated) from 185 TCGA PCa samples and 10 normal samples. The methylation levels of these MDGs were displayed using a heatmap (Fig. 1B). An intersection analysis between DEGs and MDGs, conducted using the jvenn tool, revealed that a majority of MDGs were also differentially expressed (Fig. 1C).
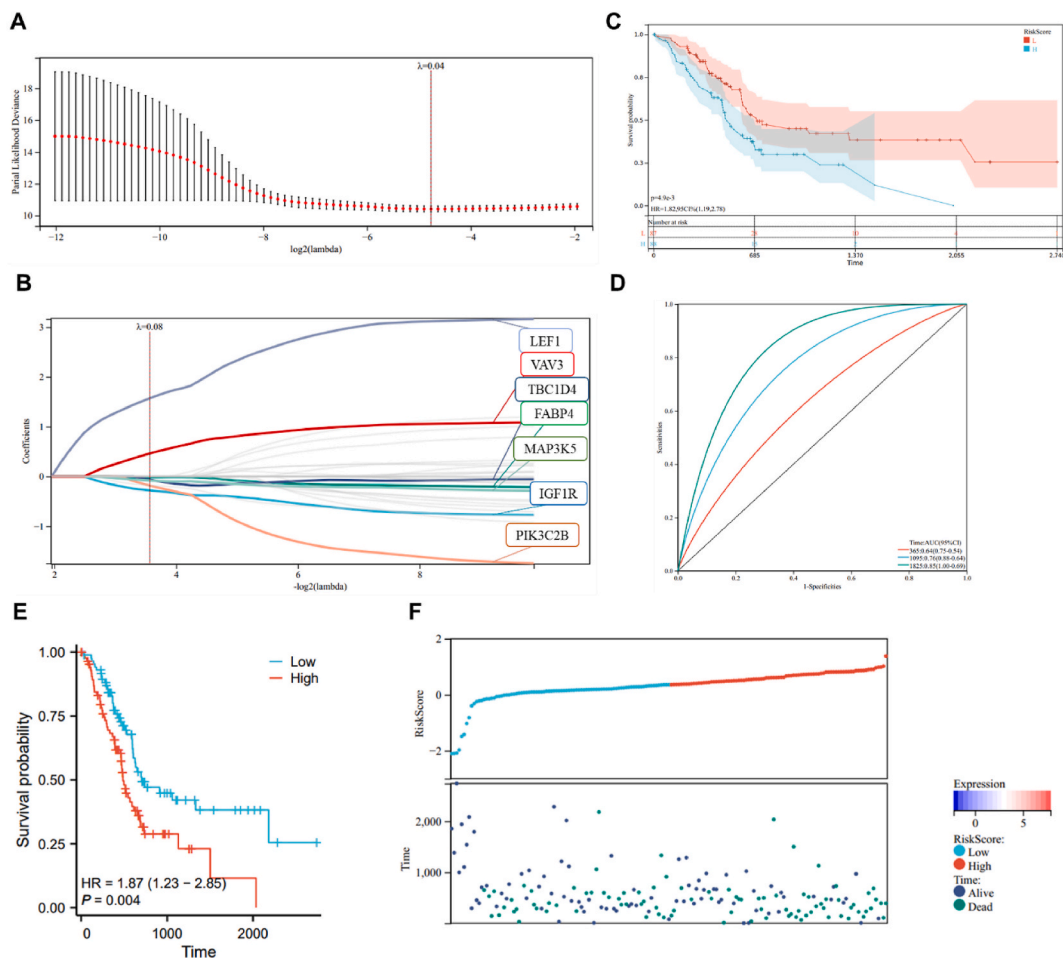

Fig. 2. Risk score model to predict prognosis of PCa based on 8 key MDGs.
Note: A: Partial likelihood deviance of LASSO coefficient distribution, Lambda = 0.04; B: Through LASSO analysis, 8 candidate genes are identified, Lambda = 0.08; C: Compare and analyze the prognostic differences between high-risk group and low-risk group of PCa patients' OS using Kaplan-Meier curve based on the risk scores analyzed by LASSO for 8 key MDGs; D: ROC analysis at 1, 3, and 5-year time points based on the risk scores analyzed by LASSO; E: Kaplan-Meier curve analysis of the progression-free survival (FPS) of PCa patients for the 8 key MDGs; F: Display the distribution of high-risk group, low-risk group, and survival status.

GO analysis revealed significant enrichment of these genes in anatomical structure morphogenesis, circulatory system development, kinase activity, protein kinase activity, and membrane raft (Fig. 1D). Pathway analysis using the KEGG database showed significant enrichment of these genes in Pathways in cancer, FoxO signaling pathway, and MAPK signaling pathway (Fig. 1E).

3.2. Prognostic risk score model based on 8 key MDGs accurately predicts the prognosis of PCa patients

To further determine the relevance of 71 MDGs to the prognosis of PCa patients, we screened 54 MDGs based on MDG-related survival data in TCGA_PADD patients by one-way Cox analysis to correlate with the prognosis of PCa patients. Next, the results of the univariate analysis were further screened using Lasso regression and multivariate Cox analysis, resulting in 8 key MDGs (LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5, PIK3C2B, IGF1R) which could be used as PCa prognostic genes (Table S3; Fig. 2A and B) and given a risk score Calculation formula. Risk score $= (-0.1231) * LEF1 + (-8e^{-04}) * ZIC3 + (-0.0732) * VAV3 + (-0.2221) * TBC1D4 + (-0.0234) * FABP4 + (0.0251) * MAP3K5 + (0.0898) * PIK3C2B + (-0.0481) * IGF1R$. Among these 8 key MDGs, the methylation levels of LEF1, ZIC3, VAV3, TBC1D4, FABP4, and IGF1R were negatively correlated with the risk score, while the methylation levels of the remaining genes were positively correlated with the risk score.

Based on the median risk score, PCa patients (excluding those with a survival time of 0 or NA, leaving a total of 175 patients) were further categorized into the high-risk group (n = 88) and the low-risk group (n = 87). Kaplan-Meier survival curves revealed a statistically significant difference in OS between the high-risk and low-risk groups (Fig. 2C). The predictive performance of the risk scoring model was assessed using the ROC curve, and the results showed that the area under the curve (AUC) for the prognostic risk scoring model based on 9 key MDGs was greater than 0.60, suggesting its accurate prediction of PCa prognosis (Fig. 2D). Furthermore, the curves of progression-free survival were analyzed for the high-risk and low-risk groups, with the high-risk group exhibiting significantly lower progression-free survival compared to the low-risk group (Fig. 2E). We also explored the relationship between the different risk scores and patient follow-up time, events, and changes in gene expression, and observed a clear decrease in patient survival rate as the risk score increased (Fig. 2F). These findings indicate that the prognostic risk scoring model based on 8 key MDGs can accurately predict PCa prognosis.

3.3. Construction of gene nomogram to predict OS prognosis in PCa patients

In order to enhance the clinical application and usability of prognostic risk scoring models, a gene-clinical-pathological nomogram was constructed by combining the genes involved in building the prognostic risk scoring model with the clinical pathological characteristics of PCa patients, including age, sex, stage classification, grade staging, and FPS. This nomogram provides a convenient tool for calculating the expected survival rate of individual PCa patients. In both univariate and multivariate regression analyses, age, PFS, and MAP3K5 were identified as significant predictive factors (Table 1). Considering all of these significant predictive factors, the constructed nomogram is presented in Fig. 3A, with a C-index of 0.68 and P < 0.001.

Moreover, the calibration curves for the OS at 1 year, 3 years, and 5 years demonstrated the accurate predictive capability of the

Table 1
Univariate and multivariate regression analysis.

| Characteristics	Univariate analysis		Multivariate analysis	
	HR (95%CI)	Pvalue	HR (95%CI)	Pvalue
Age	1.029 (1.008–1.051)	0.007	1.028 (1.001–1.055)	0.045
Sex				
MALE	Reference		Reference	
FEMALE	1.219 (0.810–1.837)	0.343	1.160 (0.667–2.019)	0.599
Stage				
StageI	Reference		Reference	
StageII	2.333 (1.069–5.090)	0.033	1.152 (0.468–2.840)	0.758
StageIV	2.100 (0.430–10.251)	0.359	1.558 (0.118–20.594)	0.737
StageIII	1.256 (0.153–10.284)	0.832	0.000 (0.000-Inf)	0.995
Grade				
G1	Reference		Reference	
G2	1.977 (1.017–3.843)	0.044	1.314 (0.533–3.241)	0.553
G3	2.614 (1.299–5.262)	0.007	1.136 (0.454–2.843)	0.785
G4	1.643 (0.210–12.824)	0.636	0.148 (0.015–1.436)	0.099
GX	1.763 (0.226–13.735)	0.588	2.779 (0.222–34.856)	0.428
PFS	0.995 (0.993–0.996)	<0.001	0.994 (0.992–0.995)	<0.001
VAV3	1.000 (1.000–1.000)	0.422	1.000 (1.000–1.000)	0.804
TBC1D4	1.000 (1.000–1.000)	0.419	1.000 (1.000–1.000)	0.474
FABP4	1.000 (1.000–1.000)	0.758	1.000 (1.000–1.000)	0.369
MAP3K5	1.000 (1.000–1.000)	0.759	1.000 (1.000–1.000)	0.448
LEF1	1.000 (0.999–1.001)	0.915	1.000 (0.998–1.002)	0.953
PIK3C2B	1.001 (1.000–1.002)	0.225	1.001 (0.999–1.002)	0.548
IGF1R	0.992 (0.986–0.997)	0.003	0.988 (0.979–0.998)	0.016
ZIC3	0.998 (0.992–1.004)	0.537	1.002 (0.995–1.009)	0.602
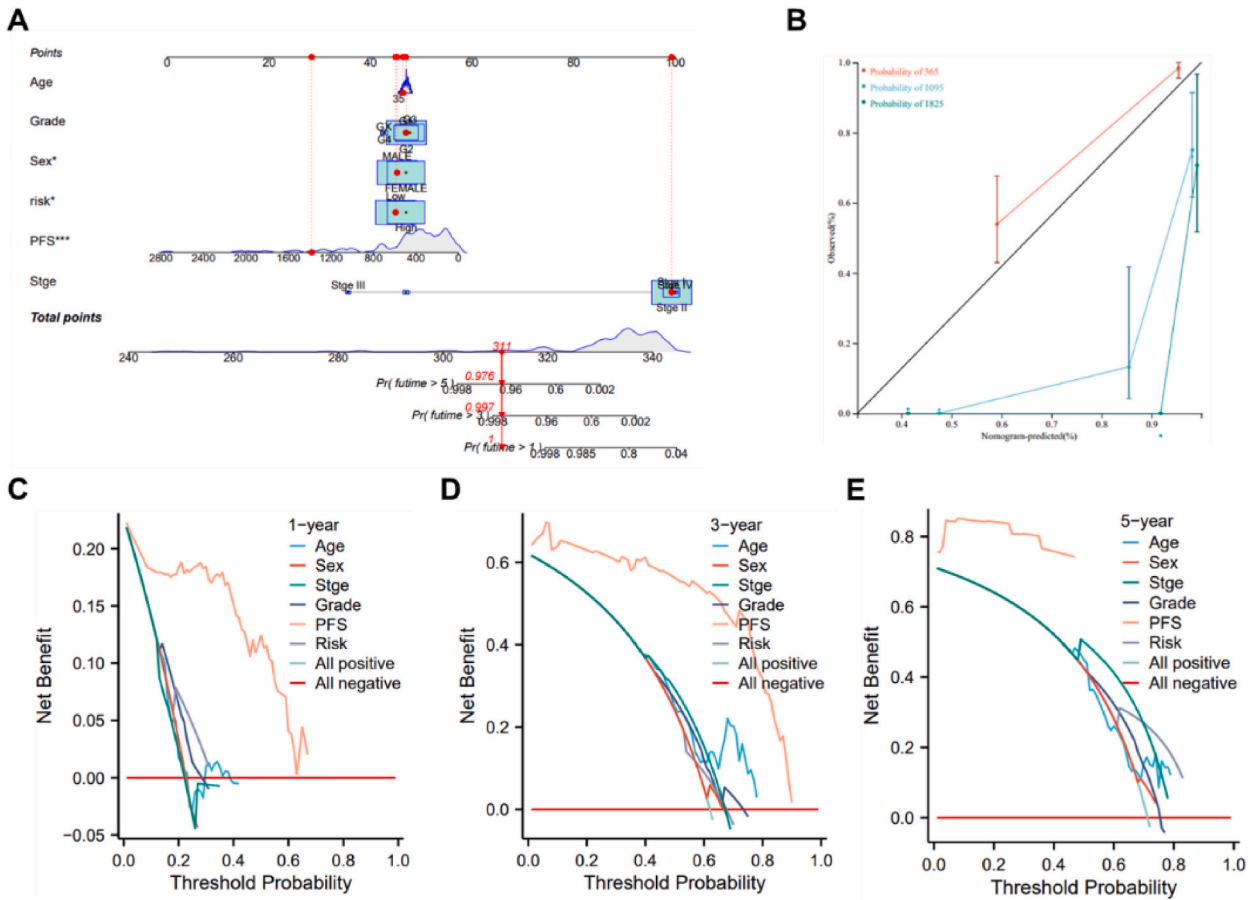
Fig. 3. Construction of gene nomogram based on 8 MDGs and clinicopathological characteristics of PCa patients.

Note: A: Plot of results of one-way Cox analysis of gene expression and clinicopathological characteristics (P value, risk factor HR, and confidence intervals are shown); B: Plot of results of multi-way Cox analysis of gene expression and clinicopathological characteristics (P value, risk factor HR and confidence intervals are shown); C: Columnar line plot constructed by MDGs (predicts 1, 3 and 5-year total survival, C_index 0.68, p < 0.001); D: OS calibration curve for the column line graph model (diagonal dashed lines indicate ideal column line graphs, blue, red and orange lines indicate observed 1-, 3- and 5-year column line graphs); E: Decision curve for OS at X years (DCA curve) (horizontal coordinates represent threshold probabilities, which could be interpreted as the probability of being grouped by a value The threshold probability could be interpreted as the number of samples greater than this value always/total number of samples after grouping by a certain value. The vertical coordinate is the net benefit, the relative benefit obtained by subtracting the proportion of true positives from the proportion of false positives weighted by the threshold probability ratio. (for other models, the model is better).

model (Fig. 3B). Additionally, the DCA revealed that the gene-based nomogram possesses strong clinical predictive ability (Fig. 3C–E). The net benefit results indicated that the combined model outperforms the single-factor model, suggesting that the nomogram can assist clinicians in accurately assessing patient prognosis. Due to its construction based on multiple prognostic factors, the nomogram is more effective than individual factors alone, performing well in short-term and long-term OS predictions. Therefore, we believe that the nomogram has the potential to aid physicians in making informed medical decisions throughout the diagnosis and treatment process of PCa.

3.4. Verification of the stability of MDG's predictions and prognosis in the external validation set

To assess the reliability and reproducibility of our prognosis model, we performed an independent external validation using the GEO dataset GSE183795 which included 139 PCa patients with complete clinical information out of a total of 134 patients. Subsequently, patients in each cohort were divided into high-risk and low-risk groups based on their stroke risk scores. Across all three validation cohorts, patients in the high-risk group exhibited poorer survival outcomes (Fig. 4A). The predictive performance of the prognosis risk scoring model constructed by MDGs was evaluated using ROC curves, which consistently showed an area under the curve (AUC) greater than 0.60, further supporting the accurate prognostication of PCa patients by our model (Fig. 4B).

Furthermore, we analyzed the relationship between different risk scores and patients' follow-up time, events, as well as the expression changes of various genes. Interestingly, as the risk score increased, the patients' survival rate significantly decreased
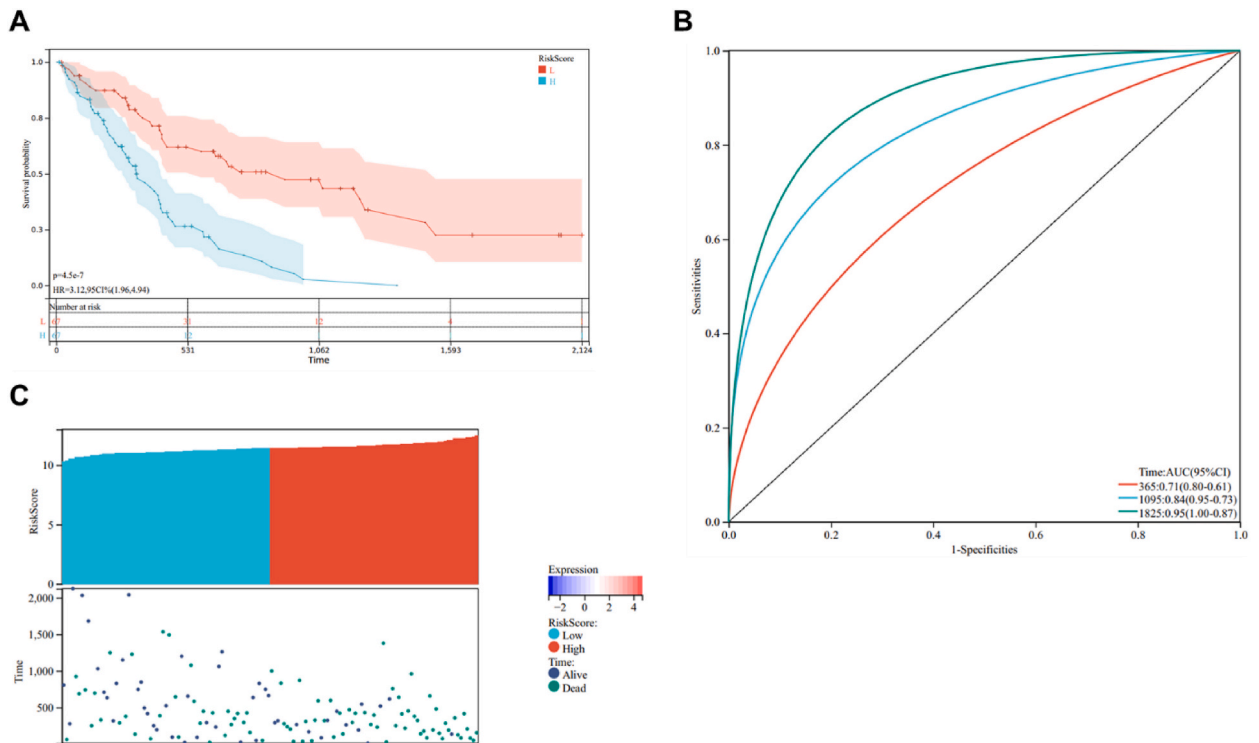
Fig. 4. Validation of the prognosis model marked by MDGs in the GEO dataset.
Note: A: Survival curve of high and low-risk groups; B: ROC curve of high and low-risk groups in the testing queue; C: Scatter plot of high and low-risk groups and survival status distribution.

(Fig. 4C).

Therefore, these results demonstrate a trend similar to the TCGA dataset and indicate the stability and reliability of the MDGs as biomarkers.

3.5. Methylation levels of 8 key MDGs in PCa tissue and normal pancreatic tissue

Of the eight key MDGs, seven (LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5 and IGF1R) were hypermethylated, while PIK3C2B was hypomethylated (Fig. 5A). In addition, there was a significant negative correlation between the methylation levels of the eight key MDGs and gene expression levels (Fig. 5B). Additionally, Kaplan-Meier survival curves were plotted based on the expression levels of the PIK3C2B gene in TCGA PCa data, revealing a negative correlation between the expression of PIK3C2B and OS of patients (Fig. 5B).

3.6. PIK3C2B is highly expressed in PCa cells, and knockdown of PIK3C2B significantly reduces the proliferation, invasion, and migration of PCa cells

The expression of LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5, IGF1R, and PIK3C2B was assessed in the human normal pancreatic cell line HPDE6c7 and PCa cell lines PANC-1, BxPC-3, and AsPC-1 using RT-qPCR. The results showed that compared to the human normal pancreatic cell line HPDE6c7, the expression levels of LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5, and IGF1R were significantly decreased, while the expression level of PIK3C2B was significantly increased in the PCa cell lines PANC-1, BxPC-3, and AsPC-1 (Fig. 6A). Additionally, we validated the expression of these eight genes in the three PCa cell lines through the CCLE website, and found that TBC1D4, MAP3K5, IGF1R, and PIK3C2B exhibited higher expression levels, while the remaining genes were expressed at lower levels (Fig. S1). Subsequently, we knocked down PIK3C2B in PANC-1 cells using shRNA, and RT-qPCR and WB results showed a significant decrease in PIK3C2B expression in PANC-1 and BxPC-3 PCa cell lines compared to the sh-NC group, with sh-PIK3C2B-1 showing a more pronounced reduction (Fig. 6B, Fig. S2A). Therefore, sh-PIK3C2B-1 was selected for subsequent experiments and referred to as sh-PIK3C2B. CCK8, scratch assay, and Transwell assay revealed a significant decrease in the proliferation, migration, and invasion abilities of PANC-1 and BxPC-3 cells in the sh-PIK3C2B group compared to the sh-NC group (Fig. 6C–E, Figs. S2B–D). Collectively, PIK3C2B is highly expressed in PCa cells, and the knockdown of PIK3C2B significantly reduces their proliferation, invasion, and migration abilities.
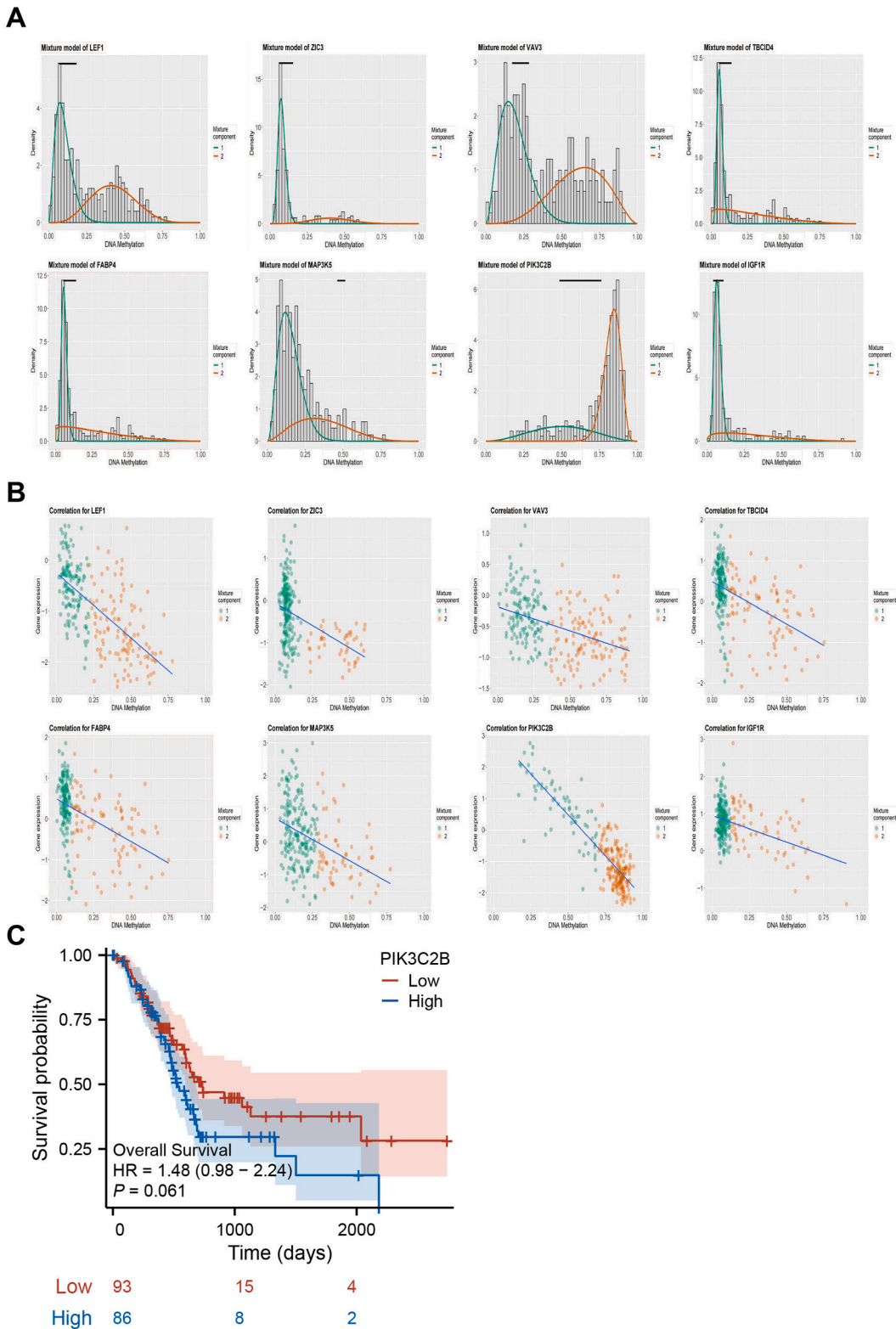
Fig. 5. Methylation level and correlation analysis of 8 key MDGs in PCa tissues and normal pancreatic tissues

Note: A: Histogram of the distribution of methylation status of MDGs (histogram shows the distribution of LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5 methylation in the samples, Beta values represent methylation levels (ranging from 0 to 1), horizontal black bars indicate the distribution of methylation values in non-tumor samples); B: Regression analysis between gene expression levels of MDGs and methylation regression analysis

between gene expression levels and methylation levels (vertical axis is gene expression levels, the horizontal axis is methylation levels). Kaplan-Meier survival curves for PIK3C2B gene expression in TCGA PCa samples, with the red curve representing samples with low PIK3C2B expression and the blue curve representing samples with high PIK3C2B expression.
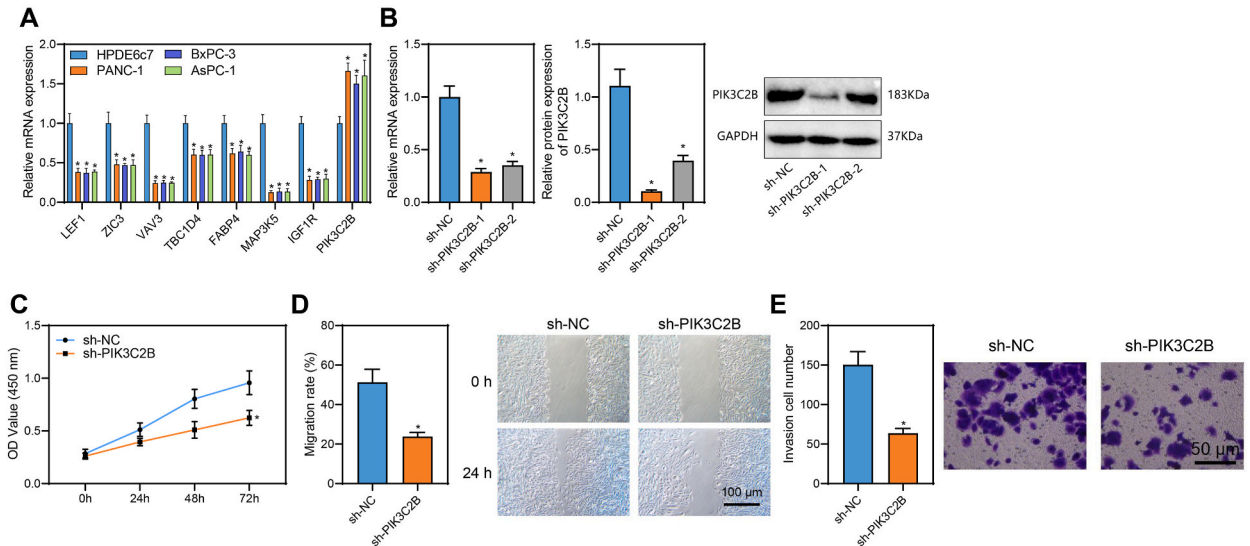

Fig. 6. Knockdown of PIK3C2B significantly inhibits the proliferation, migration, and invasion of PCa cell lines PANC-1
A: RT-qPCR to detect the expression of LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5, IGF1R, and PIK3C2B in the human normal pancreatic cell line HPDE6c7 and PCa cell line PANC-1, BxPC-3 and AsPC-1; B: RT-qPCR and Western Blot to detect the expression of PIK3C2B in the PANC-1 cell line after shRNA knockdown of PIK3C2B PIK3C2B; C: CCK8 assay to detect changes in the proliferative capacity of PANC-1 cell lines after knockdown of PIK3C2B; D: Scratch assay to detect changes in migratory capacity of PANC-1 cell lines after knockdown of PIK3C2B cell lines; E: Transwell assay to detect changes in invasive capacity of PANC-1 cell lines after knockdown of PIK3C2B cell lines changes in the invasive ability of PANC-1 cell lines after knockdown of PIK3C2B cell lines. Cell experiments were performed in three independent biological replicates. Compared with the human normal pancreatic cell line HPDE6c7 or sh-NC group, * indicates $p < 0.05$.

4. Discussion

We screened 2737 DEGs and 71 MDGs in PCa samples by bioinformatics analysis. Further, we screened the results of univariate analysis using Lasso regression and multivariate Cox analysis to finally obtain 8 key MDGs (LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5, PIK3C2B, IGF1R). These results provide some theoretical basis for subsequent PCa-related studies, and the eight key MDGs could establish risk-scoring models that could improve the prediction of the prognosis of PCa patients. LEF1 is a downstream mediator of the Wnt/β-catenin signaling pathway that maintains stem cell and organ development, and abnormal expression is associated with cancer cell proliferation, migration, and invasion. Therefore, it could be a prognostic biomarker [28]. In addition, low expression of LEF1 has been found to promote carcinogenesis and progression [29]. ZIC3 is a transcription factor with varying functions in gastric and lung cancers, and LncRNA IGBP1-AS1 has been shown to reduce ZIC3 expression and promote proliferation and invasion in breast cancer [30]. TBC1D4 may be a cancer suppressor gene, acting through the sponge miR-21 in neuroblastoma [31]. MAP3K5 is an important early responder gene in the P38MAPK pathway and an apoptosis-related gene. MAP3K5 was significantly under-expressed in tumor tissue [32]. Low expression of CircRNF13 accelerated IGF1R-induced malignant progression of PCa by targeting miR-139-5p and inhibiting IGF1R expression, thereby promoting proliferation, migration, and invasion of PCa [33]. In addition, elevated levels of PIK3C2B expression may lead to enhanced tumor cell invasion [34].

The Kaplan-Meier survival curves showed a statistically significant difference in the OS of PCa patients in the high-risk and low-risk groups. In addition, the ROC curves confirmed that the prognostic risk scoring model based on the eight key MDGs could accurately predict the prognosis of PCa patients. Methylation inhibits gene transcription, leads to gene inactivation, and, in most cases, promotes tumor progression. Our study constructed a risk assessment model based on eight key MDGs that better integrate key genes associated with PCa progression. The model may be of great importance in providing clinicians with a tissue-based prognostic assessment tool to aid their treatment decisions, which is expected to improve patient survival. Targeted methylation profiling of tumor tissue and adjacent normal tissue from 50 patients with stage I-II PCa has been used to construct and validate a tissue-based DNA methylation risk score model to predict prognosis and identify PCa patients at high risk of mortality at the surgery with a significant correlation between risk score and OS [35]. Other investigators have also developed a combined predictive risk assessment model consisting of three m6A methylation regulators (i.e., IGF2BP2, IGF2BP3, and METTL16) and four iron deposition regulators (i.e., ENPP2, ATP6V1G2, ITGB4, and PROM2) that more effectively reflects PCa progression and prognosis [36].

We combined the genes in constructing the prognostic risk score model with the clinicopathological characteristics (age, gender, pT

stage, pN stage, pM stage) of PCa patients to construct a gene nomogram. DCA showed that the gene nomogram had strong clinical predictive power. This result indicates that the accuracy of our prognostic risk assessment model is very high, and the expression abundance and methylation of key driver genes, which could correspond to the clinicopathological characteristics of PCa patients themselves, especially the clinical staging analysis of patients' disease, could accurately predict the prognosis of PCa patients. One investigator analyzed prognostic factors to create a labeled graph to predict overall OS in patients with advanced pancreatic body tail cancer (PBTC, a type of PCa), using univariate and multivariate Cox regression analysis on 1256 patients with advanced PBTC, identifying age, grade, stage N, stage M, surgery and chemotherapy as independent risk factors, demonstrating in both calibration and DCA curves that the model in the training and validation cohorts Age, grade, M stage, and surgery were independently associated with OS in good agreement in the calibration and DCA curves [37]. Other researchers showed that age, histologic subtype, primary site, N stage, and tissue metastasis were risk factors for the development of PCLM through multivariate logistic regression analysis based on the SEER database; models were developed to correlate prognostic factors such as age, grade, tumor size, histologic subtype, surgery, chemotherapy, and tissue metastasis with patient OS. DCA showed that both clinical practice prediction models were highly reliable [38]. In a report with the highest relevance to our findings, researchers developed a novel model of prognosis and drug sensitivity consisting of three DNA repair-related genes, DRG (RECQL, POLQ, and RAD17), with DCA curves and calibration curves showing that risk scores and TNM staging coincide, could be used to predict prognosis and treatment response in PCa [39].

We also found that among the 8 key MDGs, 7 (LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5, and IGF1R) were hypermethylated, while PIK3C2B was hypermethylated. 8 key MDGs were not only affected by methylation but also by gene amplification, deletion, and mutation. In addition, GSEA pathway enrichment analysis revealed that the 8 key MDGs might be involved in various processes associated with PCa, such as immune escape, inflammatory response, apoptosis, and cell migration. These results have a very important clinical guidance value, and the eight key MDGs have the potential to be markers or drug targets for early diagnosis and therapeutic options to assess PCa. Our findings provide the necessary bioinformatics and relevant theoretical basis for future PCa research, potential biomarkers for precision treatment and prognosis prediction, and a basis for future research into the molecular mechanisms of PCa. Relevant to the findings are reports that LEF1 has strong and diffuse nuclear markers in solid pseudopapillary tumors of the pancreas, is highly methylated, and could be used as a diagnostic marker [40]. In addition, inhibition of IGF1R expression in PCa cells has been reported to promote proliferation, migration, and invasion of PCa, accelerating IGF1R-induced malignant progression of PCa [33]. PIK3C2B overexpressed in tumor cells inhibits apoptosis and leads to chemoresistance in tumor cells [41].

5. Conclusions

In summary, we could draw the following preliminary conclusions: whole-genome DNA methylation sequencing technology screened eight key MDGs (LEF1, ZIC3, VAV3, TBC1D4, FABP4, MAP3K5, PIK3C2B, IGF1R) as prognostic markers for PCa patients, and the prognostic risk prediction model constructed based on the eight key MDGs could accurately predict the prognosis of PCa patients (Fig. S3). Furthermore, the prognostic risk prediction model based on the eight key MDGs could accurately predict the prognosis of PCa patients (Fig. S3). In addition, these eight key MDGs may be involved in various processes associated with PCa, such as immune escape, inflammatory response, apoptosis, and cell migration. Therefore, these genes could be markers for early diagnosis, therapeutic options, or drug targets to assess PCa. Although further experimental validation is needed, our findings provide the necessary bioinformatics and relevant theoretical basis for future PCa research. Our study provides potential biomarkers for precise treatment and prognosis prediction by analyzing MDGs in PCa. Furthermore, it provides a basis for future research into the molecular mechanisms of PCa.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability statement

The data that supports the findings of this study are available on request from the corresponding author.

CRediT authorship contribution statement

Chao Song: Writing – review & editing, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Ganggang Wang:** Writing – review & editing, Writing – original draft, Data curation. **Mengmeng Liu:** Writing – review & editing, Writing – original draft, Investigation, Data curation, Conceptualization. **Zijin Xu:** Writing – original draft, Software, Data curation, Conceptualization. **Xin Liang:** Writing – original draft, Validation, Data curation. **Kai Ding:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Yu Chen:** Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Wenquan Wang:** Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Wenhui Lou:** Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Liang Liu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e29914.

References

[1] L. Rahib, B.D. Smith, R. Aizenberg, A.B. Rosenzweig, J.M. Fleshman, L.M. Matrisian, Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States, Cancer Res. 74 (11) (2014) 2913–2921.
[2] M.C.S. Wong, J.Y. Jiang, M. Liang, Y. Fang, M.S. Yeung, J.J.Y. Sung, Global temporal patterns of pancreatic cancer and association with socioeconomic development, Sci. Rep. 7 (1) (2017) 3165.
[3] A. McGuigan, P. Kelly, R.C. Turkington, C. Jones, H.G. Coleman, R.S. McCain, Pancreatic cancer: a review of clinical diagnosis, epidemiology, treatment and outcomes, World J. Gastroenterol. 24 (43) (2018) 4846–4861.
[4] Y. Guo, W. Liu, J. Wu, Helicobacter pylori infection and pancreatic cancer risk: a meta-analysis, J. Cancer Res. Therapeut. 12 (Supplement) (2016) C229–C232.
[5] M. Umer, Z. Herceg, Deciphering the epigenetic code: an overview of DNA methylation analysis methods, Antioxidants Redox Signal. 18 (15) (2013) 1972–1986.
[6] S.D. Henriksen, P.H. Madsen, A.C. Larsen, M.B. Johansen, A.M. Drewes, I.S. Pedersen, H. Krarup, O. Thorlacius-Ussing, Cell-free DNA promoter hypermethylation in plasma as a diagnostic marker for pancreatic adenocarcinoma, Clin. Epigenet. 8 (2016) 117.
[7] J.M. Yi, A.A. Guzzetta, V.J. Bailey, S.R. Downing, L. Van Neste, K.B. Chiappinelli, B.P. Keeley, A. Stark, A. Herrera, C. Wolfgang, E.P. Pappou, C.A. Iacobuzio-Donahue, M.G. Goggins, J.G. Herman, T.H. Wang, S.B. Baylin, N. Ahuja, Novel methylation biomarker panel for the early detection of pancreatic cancer, Clin. Cancer Res. 19 (23) (2013) 6544–6555.
[8] G.P. Pfeifer, Defining driver DNA methylation changes in human cancer, Int. J. Mol. Sci. 19 (4) (2018).
[9] M. Koutsioumpa, M. Hatziapostolou, C. Polytarchou, E.J. Tolosa, L.L. Almada, S. Mahurkar-Joshi, J. Williams, A.B. Tirado-Rodriguez, S. Huerta-Yepez, D. Karavias, H. Kourea, G.A. Poultsides, K. Struhl, D.W. Dawson, T.R. Donahue, M.E. Fernandez-Zapico, D. Iliopoulos, Lysine methyltransferase 2D regulates pancreatic carcinogenesis through metabolic reprogramming, Gut 68 (7) (2019) 1271–1286.
[10] N. Sato, N. Fukushima, R.H. Hruban, M. Goggins, CpG island methylation profile of pancreatic intraepithelial neoplasia, Mod. Pathol. 21 (3) (2008) 238–244.
[11] J.D. Mizrahi, R. Surana, J.W. Valle, R.T. Shroff, Pancreatic cancer, Lancet 395 (10242) (2020) 2008–2020.
[12] G. Chen, J. Long, R. Zhu, G. Yang, J. Qiu, F. Zhao, Y. Liu, J. Tao, T. Zhang, Y. Zhao, Identification and validation of constructing the prognostic model with four DNA methylation-driven genes in pancreatic cancer, Front. Cell Dev. Biol. 9 (2021) 709669.
[13] M. Zhu, M. Ye, J. Wang, L. Ye, M. Jin, Construction of potential miRNA-mRNA regulatory network in COPD plasma by bioinformatics analysis, Int. J. Chronic Obstr. Pulm. Dis. 15 (2020) 2135–2145.
[14] H. Pan, C. Guo, J. Pan, D. Guo, S. Song, Y. Zhou, D. Xu, Construction of a competitive endogenous RNA network and identification of potential regulatory Axis in gastric cancer, Front. Oncol. 9 (2019) 912.
[15] A. Zbinden, M. Perez-Berlanga, P. De Rossi, M. Polymenidou, Phase separation and neurodegenerative diseases: a disturbance in the force, Dev. Cell 55 (1) (2020) 45–68.
[16] S. Lu, J. Wang, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, M. Yang, D. Zhang, C. Zheng, C.J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: the conserved domain database in 2020, Nucleic Acids Res. 48 (D1) (2020) D265–D268.
[17] V.P. Balachandran, M. Gonen, J.J. Smith, R.P. DeMatteo, Nomograms in oncology: more than meets the eye, Lancet Oncol. 16 (4) (2015) e173–e180.
[18] B. Dong, Y. Chen, G. Lyu, Prognostic nomograms for predicting overall survival and cancer-specific survival of patients with very early-onset colorectal cancer: a population-based analysis, Bosn. J. Basic Med. Sci. 22 (5) (2022) 803–817.
[19] J. Gao, B.A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S.O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal, Sci. Signal. 6 (269) (2013) pl1.
[20] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C.J. Byrne, M.L. Heuer, E. Larsson, Y. Antipin, B. Reva, A.P. Goldberg, C. Sander, N. Schultz, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, Cancer Discov. 2 (5) (2012) 401–404.
[21] R.K. Powers, A. Goodspeed, H. Pielke-Lombardo, A.C. Tan, J.C. Costello, GSEA-InContext: identifying novel and common patterns in expression experiments, Bioinformatics 34 (13) (2018) i555–i564.
[22] S. Canzler, J. Hackermuller, multiGSEA: a GSEA-based pathway enrichment analysis for multi-omics data, BMC Bioinf. 21 (1) (2020) 561.
[23] C.W. Luo, M.F. Hou, C.W. Huang, C.C. Wu, F. Ou-Yang, Q.L. Li, C.C. Wu, M.R. Pan, The CDK6-c-Jun-Sp1-MMP-2 axis as a biomarker and therapeutic target for triple-negative breast cancer, Am. J. Cancer Res. 10 (12) (2020) 4325–4341.
[24] M. Ye, R. Han, J. Shi, X. Wang, A.Z. Zhao, F. Li, H. Chen, Cellular apoptosis susceptibility protein (CAS) suppresses the proliferation of breast cancer cells by upregulated cyp24a1, Med. Oncol. 37 (5) (2020) 43.

[25] J. Liu, J. Jiang, X. Hui, W. Wang, D. Fang, L. Ding, Mir-758-5p suppresses glioblastoma proliferation, migration and invasion by targeting ZBTB20, Cell. Physiol. Biochem. 48 (5) (2018) 2074–2083, https://doi.org/10.1159/000492545.

[26] S. Tao, H. Li, X. Ma, Y. Ma, J. He, Y. Gao, J. Li, Elevating microRNA-1-3p shuttled by cancer-associated fibroblasts-derived extracellular vesicles suppresses breast cancer progression and metastasis by inhibiting GLIS1, Cancer Gene Ther. 28 (6) (2021) 634–648.

[27] R. Lin, Y. Wang, Q. Chen, Z. Liu, S. Xiao, B. Wang, B. Shi, TRPM2 promotes the proliferation and invasion of pancreatic ductal adenocarcinoma, Mol. Med. Rep. 17 (6) (2018) 7537–7544.

[28] L. Santiago, G. Daniels, D. Wang, F.M. Deng, P. Lee, Wnt signaling pathway protein LEF1 in cancer, as a biomarker for prognosis and a target for treatment, Am. J. Cancer Res. 7 (6) (2017) 1389–1406.

[29] X. Chen, J. Tu, C. Liu, L. Wang, X. Yuan, MicroRNA-621 functions as a metastasis suppressor in colorectal cancer by directly targeting LEF1 and suppressing Wnt/beta-catenin signaling, Life Sci. 308 (2022) 120941.

[30] D. Chen, Y. Fan, F. Wan, LncRNA IGBP1-AS1/miR-24-1/ZIC3 loop regulates the proliferation and invasion ability in breast cancer, Cancer Cell Int. 20 (2020) 153.

[31] W. Lin, Z. Wang, J. Wang, H. Yan, Q. Han, W. Yao, K. Li, circRNA-TBC1D4, circRNA-NAALAD2 and circRNA-TGFBR3: selected key circRNAs in neuroblastoma and their associations with clinical features, Cancer Manag. Res. 13 (2021) 4271–4281.

[32] R. Chen, M. Zhang, Q. Li, H. Xiong, S. Liu, W. Fang, Q. Zhang, Z. Liu, X. Xu, Q. Jiang, The Epstein-Barr Virus-encoded miR-BART22 targets MAP3K5 to promote host cell proliferative and invasive abilities in nasopharyngeal carcinoma, J. Cancer 8 (2) (2017) 305–313.

[33] X. Liu, L. Zhou, Y. Chen, X. Jiang, J. Jiang, CircRNF13 promotes the malignant progression of pancreatic cancer through targeting miR-139-5p/igf1r Axis, JAMA Oncol. 2021 (2021) 6945046.

[34] Z. Liu, X. Li, J. Ma, D. Li, H. Ju, Y. Liu, Y. Chen, X. He, Y. Zhu, Integrative analysis of the IQ motif-containing GTPase-activating protein family indicates that the IQGAP3-PIK3C2B Axis promotes invasion in colon cancer, OncoTargets Ther. 13 (2020) 8299–8311.

[35] J. Yang, Y. Tang, J. Wang, C. Yu, H. Li, B. Yi, Y. Li, J. Zhou, Z. Zhang, Development and validation of a tissue-based DNA methylation risk-score model to predict the prognosis of surgically resected pancreatic cancer patients, Gland Surg. 11 (10) (2022) 1697–1711.

[36] T. Wu, T.Y. Qian, R.J. Lin, D.D. Jin, X.B. Xu, M.X. Huang, J. Ji, F. Jiang, L.L. Pan, L. Luo, Y.F. Ji, Q.L. Chen, M.B. Xiao, Construction and validation of a m6A RNA methylation and ferroptosis-related prognostic model for pancreatic cancer by integrated bioinformatics analysis, J. Gastrointest. Oncol. 13 (5) (2022) 2553–2564.

[37] H. Shi, Z. Chen, S. Dong, R. He, Y. Du, Z. Qin, W. Zhou, A nomogram for predicting survival in patients with advanced (stage III/IV) pancreatic body tail cancer: a SEER-based study, BMC Gastroenterol. 22 (1) (2022) 279.

[38] W. Zhang, L. Ji, X. Zhong, S. Zhu, Y. Zhang, M. Ge, Y. Kang, Q. Bi, Two novel nomograms predicting the risk and prognosis of pancreatic cancer patients with lung metastases: a population-based study, Front. Public Health 10 (2022) 884349.

[39] J. Lai, W. Chen, A. Zhao, J. Huang, Determination of a DNA repair-related gene signature with potential implications for prognosis and therapeutic response in pancreatic adenocarcinoma, Front. Oncol. 12 (2022) 939891.

[40] A.D. Singhi, M. Lilo, R.H. Hruban, K.L. Cressman, K. Fuhrer, R.R. Seethala, Overexpression of lymphoid enhancer-binding factor 1 (LEF1) in solid-pseudopapillary neoplasms of the pancreas, Mod. Pathol. 27 (10) (2014) 1355–1363.

[41] Z. Liu, C. Sun, Y. Zhang, Z. Ji, G. Yang, Phosphatidylinositol 3-kinase-C2beta inhibits cisplatin-mediated apoptosis via the Akt pathway in oesophageal squamous cell carcinoma, J. Int. Med. Res. 39 (4) (2011) 1319–1332.